

Claims

1. A method of data analysis, comprising,
 - (a) employing data comprising, a plurality of records, each of said records having an associated plurality of attributes, said plurality of records being divisible into at least two categories,
 - (b) assigning as a class of one dimensional vectors one of, said records and said attributes,
 - (c) selecting an integer value K, where K is a maximum number of uncorrelated vectors to be identified from said class of one-dimensional vectors,
 - (d) selecting a threshold value,
 - (e) choosing a first vector from said class of one-dimensional vectors as a member of an uncorrelated set of vectors, and
 - (f) performing iteratively until substantially all vectors in said class of one-dimensional vectors have been analyzed,
 - (1) selecting an additional vector from said class of vectors,
 - (2) computing a correlation parameter using said first vector and said additional vector,
 - (3) comparing said correlation parameter to said threshold value, and
 - (4) adding said additional vector to said uncorrelated set of vectors if said correlation parameter is not greater than said threshold value.

2. The method of claim 1 further comprising, determining, from said uncorrelated set of vectors, a result-effective subset of attributes that is sufficient to divide said records into said at least two categories.

3. The method of claim 1, further comprising,
- (g) determining whether there are more than K vectors in said set of uncorrelated vectors,
 - (h) if there are more than K vectors in said set,

- (1) repeating an integer N number of times steps (d), (e) and (f)(1) through (f)(4),
- (2) determining N sets of vectors that are uncorrelated,
- (3) determining whether any of said N subsets have less than or equal to K vectors, and
- (4) in response to such a determination, employing one of said N subsets having less than or equal to K vectors to determine a result-effective subset of attributes that is sufficient to divide said records into said at least two categories.

4. The method of claim 3 wherein N is 10.

5. The method of claim 3, further comprising,
 (i) upon a determination that none of said N subsets has less than or equal to K members, reducing said threshold value and repeating steps (e) through (f)(4).

6. The method of claim 1, wherein said records represent cells and said attributes are properties of said cells.

7. The method of claim 1, wherein said records represent mammals and said attributes are characteristics of said mammals.

8. The method of claim 1, wherein said records represent a sample from a mammal and said attributes are biological markers.

9. The method of claim 8, wherein said biological marker is a gene product.

10. The method of claim 8, wherein said biological marker is at least one of a protein and an mRNA.

1 11. The method of claim 1, wherein at least one of said at least two categories represents a
2 predisposition to contract a disease.

1 12. The method of claim 11, wherein said disease is leukemia.

1 13. The method of claim 1, wherein at least one of said at least two categories represents a
2 predisposition to a medical treatment efficacy.

1 14. The method of claim 1, wherein a first category represents a mammal having a first
2 phenotype and a second category represents a mammal having a second, different phenotype.

1 15. The method of claim 14, wherein the first phenotype is a disease affected phenotype.

1 16. The method of claim 14, wherein the second phenotype is a non-disease affected
2 phenotype.

1 17. The method of claim 15, wherein the disease is a cancer.

1 18. A system of data analysis, comprising,
2 a processor adapted for,

3 (a) employing data comprising, a plurality of records, each of said records having an
4 associated plurality of attributes, said plurality of records being divisible into at least two
5 categories,

6 (b) assigning as a class of one dimensional vectors one of, said records and said
7 attributes,

8 (c) selecting an integer value K, where K is a maximum number of uncorrelated
9 vectors to be identified from said class of one-dimensional vectors,

10 (d) selecting a threshold value,

11 (e) choosing a first vector from said class of one-dimensional vectors as a member of
12 an uncorrelated set of vectors,

13 (f) performing iteratively until substantially all vectors in said class of one-
14 dimensional vectors have been analyzed,

15 (1) selecting an additional vector from said class of vectors,

16 (2) computing a correlation parameter using said first vector and said
17 additional vector,

18 (3) comparing said correlation parameter to said threshold value, and

19 (4) adding said additional vector to said uncorrelated set of vectors if said
20 correlation parameter is not greater than said threshold value.

1 19. The system of claim 18, wherein said processor is further adapted for determining, for
2 said uncorrelated set of vectors, a result-effective subset of attributes that is sufficient to divide
3 said records into said at least two categories.

1 20. The system of claim 19, wherein said processor is further adapted for,

2 (g) determining whether there are more than K vectors in said set of uncorrelated
3 vectors, and

4 (h) in response to there being more than K vectors in said set,

5 (1) repeating an integer N number of times steps (d), (e) and (f)(1) through
6 (f)(4),

7 (2) determining N sets of vectors that are uncorrelated,

8 (3) determining whether there are K or fewer vectors in any of said N sets, so
9 as to determine an uncorrelated set of vectors having no more than K
10 members, and

11 (4) in response to such a determination, employing one of said N subsets
12 having less than or equal to K vectors to determine a result-effective
13 subset of attributes that is sufficient to divide said records into said at least
14 two categories.

1 21. The system of claim 20 wherein N is 10.

1 22. The system of claim 20, wherein said processor is further adapted for,
2 (i) reducing said threshold value and repeating steps (e) through (f)(4), upon a
3 determination that none of said N subsets has less than or equal to no set of uncorrelated
4 vectors has no more than K members.

1 23. The system of claim 18, wherein said records represent cells and said attributes are
2 properties of said cells.

1 24. The system of claim 18, wherein said records represent mammals and said attributes are
2 characteristics of said mammals.

1 25. The system of claim 18, wherein said records represent a sample from a mammal and said
2 attributes are biological markers.

1 26. The system of claim 25, wherein said biological marker is a gene product.

1 27. The system of claim 25, wherein said biological marker is at least one of a protein and an
2 mRNA.

1 28. The system of claim 18, wherein at least one of said at least two categories represents a
2 predisposition to contract a disease.

1 29. The system of claim 28, wherein said disease is leukemia.

1 30. The system of claim 18, wherein at least one of said at least two categories represents a
2 predisposition to a medical treatment efficacy.

1 31. The system of claim 18, wherein a first category represents a mammal having a first
2 phenotype and a second category represents a mammal having a second, different phenotype.

1 32. The system of claim 31, wherein the first phenotype is a disease affected phenotype.

1 33. The system of claim 31, wherein the second phenotype is a non-disease affected
2 phenotype.

1 34. The system of claim 32, wherein the disease is a cancer.

1 35. A computer program recorded on a computer-readable medium for graphical data
2 analysis, said computer program when operating performing said steps of,

3 (a) employing data comprising, a plurality of records, each of said records having an
4 associated plurality of attributes, said plurality of records being divisible into at least two
5 categories,

6 (b) assigning as a class of one dimensional vectors a selected one of said records and
7 said attributes,

8 (c) selecting an integer value K, where K is a maximum number of uncorrelated
9 vectors to be identified from said class of one-dimensional vectors,

10 (d) selecting a threshold value,

11 (e) choosing a first vector from said class of one-dimensional vectors as a member of
12 an uncorrelated set of vectors, and

13 (f) performing iteratively until substantially all vectors in said class of one-
14 dimensional vectors have been analyzed,

15 (5) selecting an additional vector from said class of vectors,

16 (6) computing a correlation parameter using said first vector and said
17 additional vector,

18 (7) comparing said correlation parameter to said threshold value, and

19 (8) adding said additional vector to said uncorrelated set of vectors if said
20 correlation parameter is not greater than said threshold value.

- 1 36. The computer program of claim 35, when operating, further comprising, determining,
- 2 from said uncorrelated set of vectors, a result-effective subset of attributes that is sufficient to
- 3 divide said records into said at least two categories.